

2020 Research and Testing: 2015 National Content Test Relationship Question Experiment Analysis Report

A New Design for the 21st Century

Issued February 9, 2017

Version 2.1

Prepared by Emily Seem, Julia Coombs



Contents

List of Tables	3
List of Figures	3
Executive Summary	5
1. Introduction.....	6
1.1 Scope.....	6
1.2 Intended Audience	6
2. Background	6
3. Methodology	9
3.1 Relationship Panels	9
3.2 Research Questions	12
3.3 Sample Strategy	12
3.4 Data Collection	13
3.5 Evaluation Measures	15
3.6 Weights and Variance Estimation.....	16
3.7 OMB Clearance	16
3.8 Schedule.....	17
4. Limitations	17
5. Results.....	18
6. Related Evaluations, Experiments, and/or Assessments	21
7. Dependencies	21
8. Conclusions and Recommendations	21
8.1 Conclusions.....	21
8.2 Recommendations	22
9. Knowledge Management Resolutions	22
10. Acknowledgements.....	22
11. References.....	22
12. Appendix: Acronyms and Abbreviations.....	24

List of Tables

Table 1. Contact Strategy Panel Design.....	14
Table 2. Item Nonresponse Rates in Control and Experimental Panels.....	18
Table 3. Relationship Response Distributions in Control and Experimental Panels	20

List of Figures

Figure 1. 2010 Relationship Version	10
Figure 2. New Relationship Version.....	11

[This page intentionally left blank]

Executive Summary

Study Overview

As an integral part of the U.S. Census Bureau's commitment to making changes to the design of the decennial census in order to meet the strategic goals and objectives for the 2020 Census, the 2015 National Content Test (NCT) served as the primary opportunity to test content on a nationally representative sample. The NCT used a 1.2 million housing unit sample to test content modifications, different contact strategies designed to optimize self-response, and different approaches to offering in-language materials. The focus of this report is on the relationship content testing that was conducted.

The NCT relationship content testing is the latest in the Census Bureau's efforts to improve the estimates of coupled households. The 1990 Census first introduced "unmarried partner" as a response category to the relationship-to-the-householder question. The 2000 and 2010 Censuses built on this work by changing the processing of the relationship question to more accurately represent same-sex couples. However, the 2000 and 2010 Censuses tabulated counts of same-sex couple households by analyzing responses to the questions of sex and relationship rather than featuring a relationship question with response categories that explicitly distinguish between same-sex and opposite-sex spouses and partners. This collection method is prone to errors from opposite-sex couples mismarking their sex, and in 2010, the Census Bureau released a set of modified state-level same-sex household estimates. The revised relationship question tested in the NCT, one that features explicit same-sex and opposite-sex spouse and partner response categories, has the potential to improve same-sex couple estimates.

Two relationship question versions were featured in the NCT: a slightly modified version of the 2010 Census question (used in the control panel) and a new version where the two couple relationship categories were expanded to distinguish between opposite-sex and same-sex relationships (used in the experimental panel). Item nonresponse rates and relationship response distributions were the primary analytical measures used.

Results

Replacing the relationship response options "husband or wife" and "unmarried partner" with "opposite-sex husband or wife," "same-sex husband or wife," "opposite-sex unmarried partner," and "same-sex unmarried partner" did not result in statistically different item nonresponse rates for the relationship question overall. In addition, no significant differences were found within each mode.

The use of the new categories was not associated with a significant difference in the distribution of coupled household categories. No differences were found when looking at distributions by mode.

1. Introduction

To meet the strategic goals and objectives for the 2020 Census, the Census Bureau must make fundamental changes to the design, implementation, and management of the decennial census. These changes must build on successes and address the challenges of the previous censuses while also balancing challenges of cost containment, quality, flexibility, innovation, and disciplined and transparent acquisition decisions and processes.

As an integral part of the Census Bureau's commitment to making changes to the design of the decennial census, the 2015 National Content Test (NCT) served as the primary opportunity to test content on a nationally representative sample. The NCT used a 1.2 million housing unit sample, including 20,000 housing units in Puerto Rico, to test content modifications, different contact strategies designed to optimize self-response, and different approaches to offering in-language materials. One of the key aspects of the NCT, and the focus of this report, was relationship content testing. The NCT tested a revised relationship question, one that distinguishes between same-sex and opposite-sex spouse and partner response categories. Interest lay in discerning which relationship question version elicited better data quality and accuracy. Results from the NCT are foundational to informing internal planning decisions and to guiding final design decisions about the relationship question for the 2020 Census.

1.1 Scope

The scope of the 2015 NCT relationship question experiment was to test a revised relationship question to provide insights into which question version enables better data quality and accuracy. Because relationship content was not part of the two NCT reinterview studies, this report only focuses on the self-response portion of the test.

1.2 Intended Audience

This report is designed to inform decennial census staff and contractors about the relationship content of the 2015 NCT. Both internal and external stakeholders may also use this document as a reference.

2. Background

The Census Bureau has worked for over a decade on improving estimates of coupled households. In the 1990 Census, the Census Bureau first provided "unmarried partner" as an answer category to the relationship to householder question. However, the responses for couples who reported as being same-sex and married were edited to reflect an opposite-sex married couple response in the 1990 Census. At that time, same-sex marriage was not legal in any state in the country. To provide a more accurate representation of same-sex married couples, rather than assuming that they were opposite-sex couples who had mismarked sex for one spouse, the Census Bureau changed the processing of the relationship question for the 2000 and 2010 Censuses. In this new processing, couples who reported as being same-sex and married were now counted as same-sex

unmarried partners. The Census Bureau also released special reports examining the effects of processing and editing on relationship (O’Connell and Gooding, 2007; O’Connell and Feliz, 2011).

Same-sex couple households are a growing group, partially because of recent gains in social and legal recognition of same-sex marriages. According to 2010 American Community Survey (ACS) data, same-sex couples comprise approximately 1 percent of all coupled households. Same-sex couple household counts vary geographically, with Wyoming exhibiting the lowest same-sex couple household percentage at 0.29, and the District of Columbia exhibiting the highest at 4.01 (Lofquist, 2011).

The 1996 Defense of Marriage Act (DOMA) prohibited federal recognition of same-sex married couples for some federal programs and benefits. This affected the degree to which federal agencies had a policy need for federal data regarding same-sex couples. In June 2013, the Supreme Court nullified the federal definition of marriage enacted by DOMA, thereby lifting any ban on federal agencies from administering federal programs to same-sex married couples. In June 2015, the Supreme Court ruled that, pursuant to the 14th Amendment of the U.S. Constitution, states must issue marriage licenses to same-sex couples and must recognize same-sex marriages performed in other states. These two Supreme Court decisions may affect the data processing procedures and the estimates published by federal agencies.

Recent data products from the ACS and the previous two decennial censuses have tabulated counts of same-sex couple households by analyzing responses to the questions of sex and relationship to householder. The two categories to describe an intimate relationship have been “husband or wife” or “unmarried partner,” with no explicit categories to delineate same-sex and opposite-sex couples. Because same-sex couples are scarce in number compared to opposite-sex couples, a small mismark rate of sex for opposite-sex couples can greatly inflate the reporting of same-sex couple households (O’Connell and Gooding, 2006). Because of implications that mismarks have for the accuracy of counts, the Census Bureau conducted research about questionnaire layout and processing and editing rules to help determine methods to best elicit improved data quality and accuracy for the relationship question.

One such research study examined ACS data from 2005 to 2008. A substantial decline in the count of same-sex households occurred between 2007 and 2008. Researchers determined that the 2008 estimates were more accurate; the sharp decline in same-sex households was explained by the new data editing and processing rules and the new questionnaire layout that debuted in the 2008 ACS. These changes resulted in a lower mismark rate of opposite-sex couples being reported as same-sex couples. As a result of their findings on how to elicit better data quality and accuracy, the researchers recommended using a sequential layout instead of a grid and arranging the categories in a horizontal fashion for questions with two response categories on future paper questionnaires – in this case, the sex question (O’Connell, et al., 2010).

A second research study looked at same-sex couple estimates from the 2010 Census, and, in particular, why same-sex spousal households were estimated at 349,000, an estimate significantly larger than the 2010 ACS estimate of 152,000. Same-sex couple household estimates were compared by mode and phase of data collection for the ACS and 2010 Census counts. A statistical names directory was used to assign a probability that the corresponding sex for a particular name was mismarked.¹ The statistical names modeling concluded that 28 percent of the 2010 Census same-sex couple households were likely mismarked opposite-sex couples. Because of this large percentage of mismarks, the Census Bureau released a set of modified state-level same-sex household estimates. Researchers identified the grid-based format of the 2010 Census Nonresponse Followup as a primary contributor to the same-sex household overcounts. In contrast, the ACS used measures to guard against overreporting, namely a horizontal layout for the sex question response categories and automated edits in telephone and personal-visit data collection phases (O’Connell and Feliz, 2011).

In 2010, the Interagency Working Group on Measuring Relationships in Federal Household Surveys was established to identify methods to improve the federal data collection and reporting of household relationships. It is convened and chaired by the Statistical and Science Policy Branch of the Office of Management and Budget (OMB) and consists of approximately 30 representatives from a variety of federal agencies. In furtherance of this mission, the Census Bureau conducted focus groups and cognitive testing to develop a revised relationship question. The focus groups revealed three common themes. First, respondents desired new categories to represent other legal unions (e.g., civil unions and domestic partnerships, since this was before same-sex marriage was widely available) for same-sex couples. Second, respondents desired for the “unmarried partner” category to be moved up in the list of relationship response choices from its position near the bottom of the relationship response options, to be treated as equivalent to the spousal category. Last, many respondents reported the belief that the term “partner” applied more to same-sex intimate relationships. Despite this, opposite-sex unmarried couples still reported being willing to select “unmarried partner” as their relationship category (Bates, et al., 2010).

As a result of these focus group findings, two alternative relationship questions were developed. Both alternative questions underwent cognitive testing. In these tests, most respondents selected their relationship response in accordance to their legal partnership status (Demaio and Bates, 2011). The question version tested in the 2015 NCT was the one that was recommended for use following this cognitive testing. The Census Bureau also tested this version in several other quantitative tests with promising results. The largest of these was the 2013 American Housing Survey (AHS), with a nationally representative sample of approximately 160,000 households. Both unit and item nonresponse rates were not significantly different between control and

¹ The names directory was created by looking at the number of times each name was reported as male or female. A value was assigned to each name to quantify the number of times out of 1,000 that each name was reported as male.

experimental panels in the 2013 AHS (Interagency Working Group on Measuring Relationships in Federal Household Surveys, 2014). The 2014 Census Test that was conducted in portions of Maryland and the District of Colombia also tested the relationship question version that was recommended following cognitive testing even though testing content was not a stated objective of the 2014 Census Test. As with the 2013 AHS, unit and item response rates largely did not differ by question type, though nonresponse to the experimental question was slightly higher than to the 2010 version on the paper form. The 2014 Census Test also found no difference in relationship distributions between the two relationship panels (Bentley and Rothhaas, 2016). The new version was also tested in other Census Bureau surveys including the ACS and the Survey of Income and Program Participation (currently used in production). It is also being phased into production in the Current Population Survey. The 2015 NCT served as a key opportunity to add to the body of evidence for the performance of the revised relationship question by repeating the experiment using a large nationally representative sample.

3. Methodology

3.1 Relationship Panels

The 2015 NCT featured two versions of the relationship question: the 2010 version and the new version. Both versions were identical to those used in previous smaller tests. The remainder of this section provides textual and pictorial descriptions of the two versions of the relationship question that were used in the NCT and briefly explains the other experiments on the content and contact strategies in the NCT.

Figure 1. 2010 Relationship Version

2. How is this person related to Person 1? Mark ☒ ONE box.

- ☐ Husband or wife
- ☐ Unmarried partner
- ☐ Biological son or daughter
- ☐ Adopted son or daughter
- ☐ Stepson or stepdaughter
- ☐ Brother or sister
- ☐ Father or mother
- ☐ Grandchild
- ☐ Parent-in-law
- ☐ Son-in-law or daughter-in-law
- ☐ Other relative
- ☐ Roomer or boarder
- ☐ Housemate or roommate
- ☐ Foster child
- ☐ Other nonrelative

2010 relationship version: Figure 1 depicts the 2010 version of the relationship question. The question stem was the same as for the 2010 Census. However, two modifications were made to the response categories. First, the “unmarried partner” category appeared directly following the “husband or wife” category to reflect respondent preferences communicated during focus groups. Second, the category “foster child,” which was omitted in the 2010 Census because of space constraints, was reintroduced. The Census Bureau strives to keep ACS and decennial census content consistent where overlap occurs, and the ACS includes “foster child” as a response category for the relationship question. The 2010 relationship version was used as the control.

Figure 2. New Relationship Version

2. How is this person related to Person 1? Mark ☒ ONE box.

- ☐ Opposite-sex husband/wife/spouse
- ☐ Opposite-sex unmarried partner
- ☐ Same-sex husband/wife/spouse
- ☐ Same-sex unmarried partner
- ☐ Biological son or daughter
- ☐ Adopted son or daughter
- ☐ Stepson or stepdaughter
- ☐ Brother or sister
- ☐ Father or mother
- ☐ Grandchild
- ☐ Parent-in-law
- ☐ Son-in-law or daughter-in-law
- ☐ Other relative
- ☐ Roomer or boarder
- ☐ Housemate or roommate
- ☐ Foster child
- ☐ Other nonrelative

New relationship version: Figure 2 depicts the new version of the relationship question. This version included the same basic response categories as in the 2010 version, but expanded the “husband or wife” and “unmarried partner” categories to distinguish between same-sex and opposite-sex relationships. The new response categories were developed based on work done by the Census Bureau spanning multiple years. This work involved conducting focus groups, cognitive testing, discussions with other countries that are already collecting data to allow for the estimation of same-sex married couple households, and consulting with a group of subject-matter experts. The new relationship version was used for the experimental panel.

During Internet data collection, respondents who saw the new relationship version were also eligible for a couple of additional questions that checked the consistency of the response to the relationship question and the responses to the sex question. If an opposite-sex spouse or partner response option was selected on the relationship question, but the sex of the householder and the sex of the spouse or partner were reported as the same, the relationship of that person to the householder and the sex of the householder and the spouse or partner were verified through additional questions. These questions were also displayed if the respondent said that a person was a same-sex spouse or partner of the householder, but the householder and the spouse or partner were not reported as the same sex. The intent of these additional questions was to correct

any mismarks that could have occurred and thereby reduce errors in the reported data. While this report does not include any measures of how often responses to the relationship question were changed at the consistency check questions, the Census Bureau has presented on the use of the consistency check questions based on preliminary results from the 2015 NCT and recommended its inclusion to reduce misclassifications (Bates, et al., 2015a; Bates, et al., 2015b; Kreider, et al., 2016a; Kreider, et al., 2016b).

3.2 Research Questions

- 1) Do item nonresponse rates differ between control and experimental panels?
- 2) Does the distribution of coupled household categories differ between control and experimental panels?

3.3 Sample Strategy

Census Bureau researchers developed a nationally representative sample of 1.2 million housing units in the United States for the 2015 NCT. The 2015 NCT sampling methodology was designed to measure content testing differences for relatively small population groups. This sample design consisted of selecting 1,180,000 households from the 50 states and the District of Columbia and 20,000 households from Puerto Rico. The sampling frame was built from the Master Address File and excluded households that were selected in the 2015 ACS and its supplements, were in the Savannah, Georgia, designated marketing area, were selected in the 2015 Census Test in Maricopa County, or had bad address values. Group quarters were also excluded.

The stateside sample design used a stratified, systematic sampling method that oversampled census tracts that were susceptible to coverage errors, contained relatively high percentages of various race and ethnic groups, and had low self-response propensities. The stateside sample of 1,180,000 households was divided into three portions: coverage, race/ethnicity, and optimizing self-response.

To sample the coverage portion, the stateside sampling universe was subset to only include the tracts that had been flagged as susceptible for coverage errors based on 2009-2013 ACS data. These subset tracts were stratified into the six special coverage groups, and then a sample of 180,000 households was selected for the coverage portion of the sample.

Next, the remaining households in the universe that were not selected for the coverage portion of the sample were stratified into one of six race strata, based on race, ancestry, and Hispanic origin data from 2010 Census data and 2009-2013 ACS data. The sample was designed to ensure that the unbiased estimates from the test accurately reflected the nation as a whole, across a variety of demographic characteristics, by oversampling various race and ethnic groups. The strata were formed of tracts containing specific proportions of Middle Eastern or North African populations, American Indian or Alaska Native populations, Asian and Native Hawaiian or Other Pacific

Islander populations, Black or African American populations, or Hispanic or Latino populations. The selection eligibility was done sequentially, with a final selection being drawn from a strata containing tracts that did not have the minimum proportion of the groups of interest.

After the coverage and race/ethnicity portions had been selected, the remaining households in the universe were stratified into three response propensity strata based on the number of high-speed Internet connections reported by the Federal Communications Commission and on the likelihood of responding to a mail survey from the Planning Database. A sample of 200,000 housing units was selected for the optimizing self-response portion of the sample. After the 1.18 million housing unit sample was selected, the multiple sampling flags were assigned, including ones that indicated which version of the mail and Internet questionnaires the household would receive as well as which contact strategy would be used for that household.

Note, for the Puerto Rico sample, only the housing units in the San Juan Municipio were eligible for the 2015 NCT. A systematic sample of 20,000 households was selected and the appropriate sampling flags were assigned.

For more information on the sample design of the stateside or Puerto Rico sample, please refer to Mathews (2015).

3.4 Data Collection

Data was collected through three modes for the NCT: Internet, telephone through the Telephone Questionnaire Assistance (TQA) operation, and mail through the use of paper questionnaires. Nine different contact strategies for optimizing self-response were tested. As such, the timing of the mailing of the paper questionnaire varied depending upon which contact strategy a household was assigned. Table 1 illustrates the mailing schedule that was followed for the different contact strategies.

Table 1. Contact Strategy Panel Design

	Panel	#1 (August 24)	#2 (August 31)	#3* (September 8)	#4* (September 15)	#5* (September 22)
1	Internet Push (Control)	Letter	Postcard	Postcard	Mail Questionnaire	
2	Internet Push with Early Postcard	Letter	Postcard (August 25)	Postcard	Mail Questionnaire	
3	Internet Push with Early Questionnaire	Letter	Postcard	Mail Questionnaire	Postcard	
4	Internet Push with Even Earlier Questionnaire	Letter	Mail Questionnaire	Postcard	Postcard	
5	Internet Choice	Mail Questionnaire	Postcard	Postcard	Mail Questionnaire	
6	Internet Push with Postcard as 3rd Reminder	Letter	Postcard	Postcard	Mail Questionnaire	Postcard
7	Internet Push Postcard	Postcard	Postcard	Letter	Mail Questionnaire	
8	Internet Push with Early Postcard and 2nd Letter Instead of Mail Q	Letter	Postcard (August 25)	Postcard	Letter	
9	Internet Push with Postcard and Email as 1st Reminder (Same time)	Letter	Postcard and Email (August 25)	Postcard	Mail Questionnaire	

* Note: Targeted only to nonrespondents.

Responding by Internet or by phone was an option for all households from the beginning of the study. TQA interviewers used a modified version of the Internet instrument to collect responses over the phone, and respondents received content identical as they would have received if they had self-responded through the Internet. Of the 1.2 million households in the NCT, half were randomly assigned to receive the new relationship version via the Internet mode and half were randomly assigned to receive the 2010 relationship version. Likewise, paper questionnaires were allocated so that roughly half contained each version of the relationship question, though the relationship question version a given household may have received on the paper questionnaire may not have matched the relationship question version received on the Internet. A noteworthy exception was the Puerto Rico portion of the sample. Households in Puerto Rico received paper questionnaires with only the new relationship version; as a consequence, data analysis only includes stateside responses.

The NCT also included two telephone reinterviews for a subsample of respondents. These reinterviews focused on race and ethnicity or on household coverage; the NCT included content experiments on the race and ethnicity questions and on the questions that build the household roster, and these experiments were fully crossed with the relationship question experiment. Neither reinterview gathered relationship data. Responses to the reinterviews are therefore not included in this analysis.

3.5 Evaluation Measures

The two primary measures for this report are item nonresponse rates and the distribution of relationship responses.

Item nonresponse rates are calculated by dividing the number of data-defined² people with no response to the relationship question by the total number of data-defined people.

$$\text{Overall item nonresponse rate} = \frac{\text{Data-defined people with no response to the relationship question}}{\text{All data-defined people}} * 100 \text{ percent}$$

Item nonresponse rates by mode are calculated in a similar way to the overall item nonresponse rate, except that the universe is limited to responses made in a particular mode.

$$\text{Internet item nonresponse rate} = \frac{\text{Data-defined people with no response to the relationship question from Internet responses}}{\text{All data-defined people from Internet responses}} * 100 \text{ percent}$$

$$\text{Phone item nonresponse rate} = \frac{\text{Data-defined people with no response to the relationship question from phone responses}}{\text{All data-defined people from phone responses}} * 100 \text{ percent}$$

$$\text{Mail item nonresponse rate} = \frac{\text{Data-defined people with no response to the relationship question from mail responses}}{\text{All data-defined people from mail responses}} * 100 \text{ percent}$$

The distribution of the relationship responses is calculated by dividing the number of people with a particular response by the total number of data-defined people for each response option.

Two types of people are given special treatment in this analysis because of the special relationship data collected for them. The first is the householder, and the second is anyone on an extended roster. The householder, or reference person, is usually the person who owns or rents the housing unit. For mail responses, this is the first person listed on the form, or Person 1. For Internet and phone responses, this is the person who is selected as the owner or renter of the housing unit, or the first person listed if more than one person owns or rents the housing unit. No relationship information is asked about the householder because all other relationships are described in reference to the householder. Since no relationship question is asked for the householder, the householder is excluded from the numerator and denominator when calculating item nonresponse.

² A person is data-defined if a valid response is provided for at least two of the following demographic items: name, age or date of birth, sex, race or Hispanic origin, or relationship.

The mail form has physical limitations to the number of people for whom complete demographic data can be collected. Though the respondent may indicate anywhere from 0 to 99 people live in the housing unit through the response to the first question, the mail form itself cannot capture that many people. In the 2015 NCT, space allowed for the collection of complete demographic information for either six or seven people; the respondent could provide the names and some demographic information for four additional people on an extended roster. The full relationship question was not available for these four additional people, but a question asking if the person was related to the householder was instead asked. Since no details about the type of relationship are collected for the people on the extended roster, they are not included in the numerator or denominator when calculating relationship distributions in this report.

3.6 Weights and Variance Estimation

Household weights were developed to account for the unequal probability of selection of housing units due to oversampling of tracts containing the various coverage, racial, ethnic, and response groups of interest. The weights were also adjusted to account for differences from nonresponse. Each sampled case was placed into a cell based on the case's eligibility for the coverage, race and ethnicity, and response propensity strata. A nonresponse adjustment factor was calculated for each cell and applied to all responding cases in the cell based on the number of cases that responded in that cell.

To account for the complex sample design of the experiment, we used stratified jackknife replication estimation. Because of software and processing limitations, we used a random groups method to create the replicates. In this method, housing units were sorted in the order they were selected and reassigned to one of 250 different groups, or replicates.

To help ensure the validity of statistical inference when making multiple statistical comparisons, when applicable, multiple comparison corrections were used to maintain the family-wise error rate at $\alpha = 0.05$. Holm-Bonferroni multiple comparison adjustments were made as appropriate to adjust for the increased possibility of erroneous conclusions when multiple comparisons adjustment procedures are used. The multiple comparison corrections reduce the possibility of identifying false-positive differences and ensure that we do not cloud our ability to form inferential conclusions.

3.7 OMB Clearance

This research project 2015 National Content Test is covered under OMB clearance number 0607-0985, which expires June 30, 2018.

3.8 Schedule

Milestone	Date
First mailout	August 24, 2015
Data collection begins	August 24, 2015
Census Day	September 1, 2015
Data collection ends	October 30, 2015
First draft of report to reviewers	August 12, 2016
Draft of report presented to DROM	September 12, 2016
Draft of report sent to PMGB	October 3, 2016
Final draft of report	October 24, 2016

4. Limitations

- The 2015 NCT was not carried out in a decennial census environment, which is difficult to replicate without large-scale advertising campaigns. Because of this and other factors such as increased public trust in the decennial census compared to a survey, the NCT likely had a lower overall self-response rate than will be seen in the 2020 Census. Additionally, there was no nonresponse follow-up operation in the NCT. If answers to relationship question for nonrespondents are markedly different from those for respondents, nonresponse bias is a limitation of the NCT that should not be ignored. Weights were developed to account for nonresponse.
- Though all sampled households were sent a unique User ID with each invitation to reply, a respondent could complete an electronic census form without using his or her User ID by providing an address. Such responses are called non-ID responses. Non-ID respondents and respondents living in Puerto Rico were excluded from the analysis in this report because all non-ID respondents and respondents living in Puerto Rico received the same relationship content. Since interest lay in comparing results across two panels, a determination of which panel performs better for non-ID respondents and respondents living in Puerto Rico could not be made. This is an inherent limitation of the study design. The results from this report, therefore, can only be generalized to respondents with an ID living in the 50 states and the District of Columbia.
- Estimated relationship distributions included in this report are not meant to be measures of the actual relationship distribution in the United States at the time of the 2015 NCT. Though weights are used to account for differential sampling probabilities and for nonresponse, this report includes only self-responders in areas with relatively high address mailability. The 2015 NCT did not include a nonresponse follow-up operation or any enumeration of areas with high concentrations of nonmailable addresses. Households who usually respond through nonresponse followup or who live

in areas with high concentrations of nonmailable addresses may have different household compositions or may answer the relationship question in a different way than those included in the 2015 NCT. In order to not affect the demographic compositions we observe in the data, the weights do not impose demographic controls.

5. Results

Next are the results of the 2015 NCT relationship question experiment. The results for the first research question, that of whether item nonresponse rates differ between control and experimental panels, are presented. In order to analyze item nonresponse rates, z-tests of proportions, with Holm-Bonferroni multiple comparison adjustments made as appropriate, were used. Table 2 displays item nonresponse rates for control and experimental panels. Note that in Table 2 and in Table 3, the combined column represents overall results and does not take into account response mode. These overall numbers are then parsed into three modes: mail, Internet, and phone, which refer to the mode in which the response was received. The following definitions will herein be used to make this distinction explicit: 1) *columns* – refers to combined, mail, Internet, and phone; and 2) *modes* – refers to just mail, Internet, and phone.

Table 2. Item Nonresponse Rates in Control and Experimental Panels

Panel	Combined	Mail	Internet	Phone
Control Panel	0.6%(0.03)	0.6%(0.05)	0.5%(0.03)	1.4%(0.13)
Experimental Panel	0.6%(0.03)	0.8%(0.05)	0.5%(0.03)	1.3%(0.16)

Source: 2015 National Content Test data. Note: Estimates are weighted with standard errors in parentheses.

Table 2 illustrates that within each column, item nonresponse rates between control and experimental panels are comparable. In fact, no statistically significant differences between control and experimental item nonresponse rates were found at the 0.05 significance level within each column. The research questions focus on differences between panels rather than across modes. Still, differences across modes are particularly noticeable in the phone mode, where item nonresponse rates for both panels appear higher than for mail or Internet modes. Modal differences are to be expected because of the design of the NCT, which allowed for responses to be made in three different modes. Though this report does not focus on differences between modes, it is interesting to hypothesize reasons as to why the phone item nonresponse rates are higher. Both versions of the relationship question tested featured a long list of response categories. When a person responded by the mail or Internet mode, they could see the whole list of response categories before them and make a selection from this list. However, in the phone mode, respondents needed to listen to the interviewer list all of the categories verbally. Not only is it more difficult to remember the response categories when listed verbally instead of on a paper or computer screen visually, but the listing of response categories verbally would also become time-consuming, especially if the relationship question was asked repeatedly for different members of the household. It seems plausible that these special challenges unique to

asking the relationship question by phone could have contributed to the higher item nonresponse rates seen in both panels. Since we know that the demographics of households that respond in each mode differ by mode (Phelan, 2016), significance testing was performed only within each column and not across columns.

Next, the results for the second research question, that of whether coupled household category distributions differ between control and experimental panels, are presented. Two-sample chi-square tests, with Holm-Bonferroni multiple comparison adjustments made as appropriate, were used to analyze relationship distributions within columns. Table 3 displays the relationship distributions for control and experimental panels.

Significance testing revealed no statistically significant differences in the overall relationship distributions between control and experimental panels within each column. When looking specifically at the distribution of coupled households, it is interesting to note that in the combined, mail, and Internet columns, the estimated percentage of respondents who report as “husband or wife” is not statistically different between control and experimental panels. Similarly, the percentage of respondents who report as “unmarried partner” is not statistically different between control and experimental panels for the combined and Internet columns.

Table 3. Relationship Response Distributions in Control and Experimental Panels

	Combined		Mail		Internet		Phone	
	Unweighted Count	Weighted Percent	Unweighted Count	Weighted Percent	Unweighted Count	Weighted Percent	Unweighted Count	Weighted Percent
Control Panel								
Total Persons	671,829	100%	147,614	100%	471,440	100%	52,775	100%
Householder	266,858	41.0%(0.07)	58,816	41.4%(0.13)	178,789	39.0%(0.07)	29,253	56.3%(0.30)
Husband or wife	140,934	21.2%(0.19)	26,910	18.4%(0.17)	104,402	22.4%(0.16)	9,622	18.7%(0.18)
Unmarried partner	14,660	2.4%(0.03)	3,667	2.7%(0.06)	10,296	2.5%(0.04)	697	1.4%(0.07)
Other relative¹	224,087	32.0%(0.10)	52,552	33.8%(0.23)	160,154	33.0%(0.10)	11,381	20.3%(0.33)
Other non-relative²	22,459	2.9%(0.07)	4,865	3.0%(0.08)	16,132	2.9%(0.09)	1,462	2.6%(0.14)
Missing	2,619	0.4%(0.02)	592	0.4%(0.03)	1,667	0.3%(0.02)	360	0.6%(0.06)
Multiple Marks	212	0.0%(0.00)	212	0.1%(0.01)	0	0.0%(N/A)	0	0.0%(N/A)
Experimental Panel								
Total Persons	658,115	100%	136,835	100%	468,804	100%	52,476	100%
Householder	261,736	41.1%(0.07)	54,279	41.1%(0.13)	178,207	39.1%(0.07)	29,250	56.3%(0.32)
Husband or wife	138,326	21.2%(0.18)	24,932	18.4%(0.18)	103,766	22.4%(0.15)	9,628	19.1%(0.21)
Opposite-Sex	137,093	21.0%(0.18)	24,673	18.2%(0.18)	102,814	22.1%(0.15)	9,606	19.0%(0.21)
Same-Sex	1,233	0.2%(0.01)	259	0.2%(0.01)	952	0.2%(0.01)	22	0.0%(0.01)
Unmarried partner	14,096	2.4%(0.03)	3,268	2.6%(0.05)	10,166	2.5%(0.04)	662	1.3%(0.06)
Opposite-Sex	12,932	2.2%(0.03)	3,067	2.5%(0.05)	9,235	2.2%(0.04)	630	1.2%(0.06)
Same-Sex	1,164	0.2%(0.01)	201	0.2%(0.01)	931	0.2%(0.01)	32	0.1%(0.02)
Other relative¹	218,953	32.0%(0.11)	48,901	34.2%(0.21)	158,795	32.8%(0.11)	11,257	20.3%(0.37)
Other non-relative²	21,986	2.9%(0.07)	4,543	3.0%(0.08)	16,060	3.0%(0.08)	1,383	2.4%(0.11)
Missing	2,794	0.4%(0.02)	688	0.5%(0.03)	1,810	0.3%(0.02)	296	0.6%(0.07)
Multiple Marks	224	0.0%(0.00)	224	0.2%(0.02)	0	0.0%(N/A)	0	0.0%(N/A)

Source: 2015 National Content Test data. Note: Estimates are weighted with standard errors in parentheses.

¹ Includes the following: biological son or daughter, adopted son or daughter, stepson or stepdaughter, brother or sister, father or mother, grandchild, parent-in-law, son-in-law or daughter-in-law, and other relative.

² Includes the following: roomer or boarder, housemate or roommate, foster child, and other non-relative.

6. Related Evaluations, Experiments, and/or Assessments

- 2015 National Content Test Study Plan: Race & Ethnicity (Jones et al., 2015)
- 2015 National Content Test Report: Optimizing Self Response (Phelan, 2016)
- 2015 National Content Test Content Research Report: Within-Household Coverage (Fink, 2016)

7. Dependencies

- The 2015 NCT relationship question experiment project is dependent on the non-ID project because the results of non-ID research will impact how self-response content options are implemented. This dependency is especially important to highlight because the results of the NCT relationship content testing cannot be generalized to non-ID respondents. Hence, results from non-ID research are vital to form the most comprehensive recommendations on the relationship question content that works best for all respondents, both ID and non-ID.
- The 2015 NCT relationship question experiment project is dependent on the iterative testing process for the research and testing phase of the planning for the 2020 Census. Results from the 2014 Census Test, in which the same two relationship question versions were tested, informed the final design decisions for the NCT.

8. Conclusions and Recommendations

8.1 Conclusions

This research was designed to answer two research questions about the use of some new response options in the relationship question:

1. Do item nonresponse rates differ between control and experimental panels?
2. Does the distribution of coupled household categories differ between control and experimental panels?

The expansion of the “husband or wife” and “unmarried partner” categories so that opposite-sex and same-sex couples may be distinguishable when choosing a response category did not result in statistically different item nonresponse rates for the relationship question. This finding holds when examining item nonresponse rates for each mode individually. As expected, the results suggest that item nonresponse may not be constant across modes.

Likewise, the inclusion of the new categories was not associated with a significant difference in the distribution of coupled household categories. The same overall percentage of coupled households was measured in both the control and experimental panels. As with the findings regarding item nonresponse rates, no difference was found when looking at distributions by mode.

Collecting detailed data about types of coupled households can improve editing procedures for demographic data after data collection as well as increase the accuracy of published results related to relationships. The inability of this report to find significant differences within each mode for item nonresponse rates and for the distribution of coupled households suggests that the collection of this detailed data will not adversely affect data quality and will enable improvements in editing procedures since inconsistency between the relationship report and sex reports for the householder and their spouse or partner will flag cases so that they can be dealt with. Without the explicit listing of the type of spouse or partner on the relationship item, the cases which have mismarked sex or relationship cannot be flagged in the editing process.

8.2 Recommendations

- Replace the single response option “husband or wife” with the two response options “opposite-sex husband or wife” and “same-sex husband or wife” in all modes
- Replace the single response option “unmarried partner” with the two response options “opposite-sex unmarried partner” and “same-sex unmarried partner” in all modes
- List the spouse and unmarried partner categories at the top of the list of answer categories
- Include relationship and sex consistency edit questions in automated instruments

9. Knowledge Management Resolutions

No Knowledge Management Recommendations.

10. Acknowledgements

This analysis report is the product of many contributors. Thanks are especially given to Rose Kreider for her subject matter expertise and to Frederic Lestina for verifying the results found in this report.

11. References

- Bates, N., DeMaio, T.J., Robins, C., and Hicks, W. (2010). “Classifying Relationship and Marital Status Among Same-Sex Couples.” Proceedings of the Survey Research Methods Section. American Statistical Association, pp. 5822-5836.
- Bates, N., Lewis, J. and Lofquist D. (2015a). “Counting Gay and Lesbian Couples: Can a New Question and Edit Check Reduce Measurement Error?” Presented at the 2015 European Survey Research Association conference, Reykjavik, Iceland.
- Bates, N., Lewis, J., Lofquist D., and Streeter, M. (2015b). “If it's so easy, why is it so hard? Measuring same-sex couples.” Presentation at the AAPOR conference, Hollywood, FL.

- Bentley, M., Rothhaas, C. (2016). "2020 Census Research and Testing: 2014 Census Test Results for Optimizing Self-Response." 2020 Census Program Memorandum Series #2016.04. U.S. Census Bureau. June 14, 2016.
- DeMaio, T., and Bates, N. (2011). "New Relationship and Marital Status Questions: A Reflection of Changes to the Social and Legal Recognition of Same-Sex Couples in the U.S." Paper presented at the Annual Meetings of the American Association for Public Opinion Research, Phoenix, AZ.
- Fink, E. (2016). "2015 National Content Test Content Research Report: Within-Household Coverage." DSSD 2020 Decennial Census Memorandum Series, Draft, U.S. Census Bureau.
- Interagency Working Group on Measuring Relationships in Federal Household Surveys (2014). "Paper 1: Measuring Same-Sex Co-Residential Relationships." Statistical Policy Working Paper.
- Jones, N., Bentley, M., Konya, S., Pratt, B., Marks, R., Coombs, J., Mathews, K., and Seem, E. (2016). "2015 National Content Test Study Plan: Race & Ethnicity." U.S. Census Bureau.
- Kreider, R., Bates, N., Lofquist, D. (2016a). "Improving Measurement of Same-Sex Couple Households in Census Bureau Surveys." Presentation at AAPOR conference, Austin, TX.
- Kreider, R., Bates, N., Lofquist, D. (2016b). "Improving Measurement of Same-Sex Couple Households in Census 2020." Presentation at the Joint Statistical Meetings, Chicago, IL.
- Lofquist, D. (2011). "Same-Sex Couple Households." American Community Survey Briefs (ACSB/10-03), U.S. Census Bureau.
- Mathews, K., (2015). "Sample Design Specifications for the 2015 National Content Test." Decennial Statistical Studies Division 2020 Decennial Census R&T Memorandum Series #R-11. U.S. Census Bureau. November 18, 2015.
- O'Connell, M., and Feliz, S. (2011). "Same-Sex Couple Households from the 2010 Census." SEHSD Working Paper, U.S. Census Bureau.
- O'Connell, M., and Gooding, G. (2006). "The Use of First Names to Evaluate Reports of Gender and Its Effect on the Distribution of Married and Unmarried Couple Households." Paper presented at the Population Association of America, Los Angeles, CA.
- O'Connell, M., Lofquist, D., Simmons, T., and Lugaila, T. (2010). "New Estimates of Same-Sex Couple Households from the American Community Survey." Paper presented at the Annual Meeting of the Population Association of America, Dallas, TX.

Phelan, J. (2016). "2015 National Content Test Optimizing Self-Response Report." U.S. Census Bureau, draft.

12. Appendix: Acronyms and Abbreviations

Abbreviation	Meaning
ACS	American Community Survey
AHS	American Housing Survey
DOMA	Defense of Marriage Act
NCT	National Content Test
OMB	Office of Management and Budget
TQA	Telephone Questionnaire Assistance